

## VIDEO AND AUDIO CONTENT ANALYSIS SYSTEM

### CROSS-REFERENCE TO RELATED APPLICATIONS

The present application claims priority from US provisional Application S/N 60/264,725, filed January 30, 2001, which is incorporated herein by reference in its entirety.

### BACKGROUND OF THE INVENTION

The ever-increasing use of video and audio in the military, law enforcement and surveillance fields has resulted in the need for an integrative system that may combine several known detecting and monitoring systems. There are several questions related to real-time and off-line analysis and processing of information regarding the existence and behavior of people and objects in a certain monitored area.

Examples of such typical questions include questions regarding presence and identification of people (e.g. Is there anybody? If so, who is he?), movement (e.g. Is there anything moving?), number of people (e.g. How many people are there?), duration of time (e.g. for how long have they stayed in the area?), identifications of sounds, content of speech, number of articles and the like.

Currently, a dedicated system having a separate infrastructure is usually installed to provide a limited solution to each of the above-mentioned questions. Non-limiting examples of these systems include a video and audio recording system such as NiceVision of Nice Systems Ltd., Ra'anana, Israel, a movement-detecting system such as Vicon8i of Vicon Motion Systems, Lake Forest, California, USA and a face-recognition system such as FaceIt system of Visionics Corp., Jersey City, New Jersey, USA.

The separate infrastructure for each application also limits the area of surveillance. For example, a face recognition system, which is connected to a single dedicated video sensor, can cover only a narrow area. Moreover, the separated applications provide only a limited and partial integration between various monitoring applications.

An integrated monitoring system may enable advanced solutions for combined and conditioned questions. An example of conditioned questions is described below. "If there is a movement, is anyone present? If someone is present, can he be identified? If he can be identified, what is he saying? If he cannot be identified, record the event."

It would be advantageous to have an integrated monitoring system for analysis and processing of video and audio signal from a plurality of sources in real-time and off-line.

## SUMMARY OF THE INVENTION

The present invention is directed to various methods and systems for analysis and processing of video and audio signals from a plurality of sources in real-time or off-line. According to some embodiments of the present invention, analysis and  
5 processing applications are dynamically installed in the processing units.

There is thus provided in accordance with some embodiments of the present invention, a system having one or more processing units, each coupled to a video or an audio sensor to receive video or audio data from the sensor, an application bank comprising content-analysis applications, and a control unit to instruct the application bank to install at least one of the applications into at least one of the processing units.

There is further provided in accordance with some embodiments of the present invention, a method comprising installing one or more content-analysis applications from an application bank into one or more video or audio processing units, the applications selected according to predetermined criteria and processing input received from one or more video or audio sensors, each coupled to a respective one of the video or audio processing units according to at least one of the installed applications.

### BRIEF DESCRIPTION OF THE DRAWINGS

The subject matter regarded as the invention is particularly pointed out and distinctly claimed in the concluding portion of the specification. The invention, however, both as to organization and method of operation, together with objects, features and advantages thereof, may best be understood by reference to the following detailed description when read with the accompanying drawings in which:

Fig. 1 is a block diagram illustration of a video and audio content analysis system according to some embodiments of the present invention;

Fig. 2 is a block diagram illustration of a distributed video and audio content analysis system according to some embodiments of the present invention;

Fig. 3 is a flow chart diagram of the operation of the system of Fig. 1 and 2 according to some embodiments of the present invention; and

Figs. 4A and 4B are block diagram illustrations of the video-processing unit of Fig. 1 and Fig. 2 according to some embodiments of the present invention;

It will be appreciated that for simplicity and clarity of illustration, elements shown in the figures have not necessarily been drawn to scale. For example, the dimensions of some of the elements may be exaggerated relative to other elements for clarity. Further, where considered appropriate, reference numerals may be repeated among the figures to indicate corresponding or analogous elements.

## DETAILED DESCRIPTION OF THE PRESENT INVENTION

In the following detailed description, numerous specific details are set forth in order to provide a thorough understanding of the invention. However, it will be understood by those of ordinary skill in the art that the present invention may be practiced without these specific details. In other instances, well-known methods, procedures, components and circuits have not been described in detail so as not to obscure the present invention.

Reference is now made to Fig. 1, which is a block diagram illustration of a video and audio content analysis system 10 according to some embodiments of the present invention. System 10 may be coupled to a surveillance system having a video and audio logging and retrieval unit such as NiceVision of Nice Systems Ltd., Ra'anana, Israel.

System 10 may comprise a plurality of video sensors 12 and a plurality of audio sensors 14. Video sensor 12 may output an analog video signal or a digital video signal. The digital signals may be in the form of data packages over Internet Protocol (IP) as their upper layer and may be transmitted over digital subscriber line (DSL), asymmetric DSL (ADSL), asynchronous transfer mode (ATM) and frame relay (FR).

Audio sensor 14 may output an analog audio signal or a digital audio signal. The digital signals may be in the form of data packages over a network, for example, an IP network, an ATM network or a FR network.

System 10 may further comprise a plurality of video-processing units 16 able to receive signals from video sensors 12 and a plurality of audio-processing units 18 able to receive signals from audio sensors 14. Video-processing units 16 may be coupled to video sensors 12 and may be located in the proximity of sensors 12 or may be located

remote from sensors 12. Alternatively, video-processing units 16 may be embedded in video sensors 12. Audio-processing units 18 may be coupled to audio sensors 14 and may be located in the proximity of sensors 14 or may be located remote from sensors 14. Alternatively, audio-processing units 18 may be embedded in audio sensors 14.

5 Video-processing unit 16 and audio-processing unit 18 may be a single integral unit.

Other types of sensors and their associated processing units may be added to system 10. Non-limiting examples of additional sensors are smoke sensors, fire sensors, motion detectors, sound detectors, presence sensors, movement sensors, volume sensors, and glass breakage sensors.

System 10 may further comprise an application bank 24 coupled to processing units 16 and 18. Application bank 24 may comprise a plurality of various content analysis applications based on video and/or audio signals processing. For example, application 25 may be a video motion-detecting application, application 26 may be a video based people-counting application, application 28 may be a face-recognition application, and application 29 may be a voice-recognition application. Additional applications may be added to application bank 24. Non-limiting examples of additional applications include conversion of speech to text, compressing the video and/or audio signal and the like.

System 10 may further comprise a database 30 and a storage media 32. Storage  
20 media 32 may receive data from processing units 16 and 18 and to store video and audio input. Non-limiting examples of storage media 32 include a computer's memory, a hard disk, a digital audio-tape, a digital video disk (DVD), an advanced intelligent tape (AIT), digital linear tape (DLT), linear tape-open (LTO), JBOD, RAID, NAS, SAN and ISCSI. Database 30 may store time, date, and other annotations relating to specific segments of  
25 recorded audio and video input. For example, an input channel associated with the

sensor from which the input was received and the location of the stored input in storage 32. The type of trigger for recording, manual or scheduled, may likewise be stored in database 30. Alternatively, the segments of recorded audio and video, preferably compressed may be also stored in database 30.

5        System 10 may further comprise a control unit 20 able to control any of elements 16, 18 and 24. At least one set of internal rules may be installed in control unit 20. Non-limiting examples of a set of rules include a set of installation rules, a set of recording rules, a set of alert rules, a set of post-alert action rules, and a set of authorization rules.

10        The set of installation rules may determine the criteria for installing applications in the processing units. The set of recording rules may determine the criteria for recording audio and video data. The set of alert rules may determine the criteria for sending alert notifications from the processing units to the control unit. The set of post-alert action rules may determine the criteria for activating or deactivating applications installed in a processing unit and the criteria for re-installing applications in the processing units.

15        Control unit 20 may command application bank 24 to install various applications in processing units 16 and 18 as required by the internal rules installed in control unit 20. The installation may vary among various processing units. For example, 20 in one video-processing unit 16, application bank 24 may install motion detection application 25 and people-counting application 26. In another video-processing unit 16, application bank 24 may install motion detection application 25 and face recognition application 28.

The installation may be altered from time to time according to instructions from a time-based scheduler (not shown) installed in control unit 20 or manually triggered by an operator as will be explained below.

System 10 may further comprise at least one client computer 40 having a display and at least one speaker (not shown) and at least one printer 42. Client computer 40 and printer 42 may be coupled to database 30, storage 32, control unit 20, and application bank 24, either by direct connection or via a network 44. Network 44 may be a local area network (LAN) or a wide area network (WAN).

The operators of system 10 may control it via client computers 40. Client computer 40 may request playing a real-time stream of video and/or audio data. Alternatively, client 40 may request playback of video and audio data stored at database 30 and/or storage 32. The playback may comprise synchronized or unsynchronized recorded data of multiple audio and/or video channels. The video may be played on the client's display and the audio may be played via the client's speakers.

Client 40 may also edit the received data and may execute off-line investigation. The term "off-line investigation" refers to the following mode of operation. Client 40 may request playback of certain video and/or audio data stored in storage 30. Client 40 may also command application bank 24 to download at least one of the applications to client 40. After receiving the application and the video and/or audio files, the application may be executed by client 40 off-line. The off-line investigation may be executed even when the specific application was not installed or enabled on the processing unit 16 or 18 coupled to the sensor 12 or 14 from which the video or audio data were recorded.

Each operator may have personal authorization to perform certain operations according to a predefined set of authorization rules installed in control unit 20. Some



operators may have authorization to alter via client 40 at least certain of the internal rules installed in control unit 20. Such alteration may include immediate activation or de-activation of an application in one of processing units 18 and 16.

Client 40 may also send queries to database 30. An example of a query may be:

5 "Which video sensors detected movement between 8:00 AM and 11:00 AM?" Client 40 may also request sending reports to printer 42.

Reference is now made to Fig. 2, which is a block diagram illustration of a video and audio content analysis system 11 according to some embodiments of the present invention. System 11 is a distributed version of system 10 of Fig. 1 and elements in common may have the same numeral references. In these embodiments, video sensors 12, which may be coupled to video processing units 16 and audio sensors 14, which may be coupled to audio processing units 18 may be located at at least two remote and separate sites.

Processing units 16 and 18 may be coupled to all the other elements (e.g. database 30, storage 32, control unit 20 and application bank 24 as well as clients 40) of system 11 via network 44. Application bank 24, control unit 20, database 30 and storage 32 may be coupled to each other via network 44, which may include several networks. However, it should be understood that the scope of the present invention is not limited to such a system and system 10 may be only partially distributed.

20 Reference is now made to Fig. 3, which is a simplified flowchart illustration of the operation of the video and audio content analysis system of Figs. 1 and 2, according to some embodiments of the present invention. In the method of Fig. 3, control unit 20 may command application bank 24 to install various applications in processing units 16 and 18 (step 100). Different applications may be installed in different units. Processing  
25 units 16 and 18 may then receive video and audio signals from video and audio sensors

12 and 14, respectively (step 102). If the signals are analog signals, processing units 16 and 18 may convert the analog signals to digital signals.

Processing units 16 and 18, then, may execute the applications installed in each unit (step 104). The audio and video signals may be compressed and stored in storage media 32 according to a predefined set of recording rules installed in control unit 20 (step 106).

Processing units 16 and 18 may also output indexing-data to be stored in database 30 (step 108). Non-limiting examples of indexing data may include the time of recording, time occurrence of matching a voice or face and the time of counting. Other non-limiting examples may include a video channel number, an audio channel number, results of a people-counting application (e.g. number of people), an identifier of the recognized voice or the recognized face and direction of movement detected by a motion detection application.

Processing unit 16 or 18 may alert control unit 20 when one of the applications installed in it detects a condition corresponding to one of the predefined alert rules (step 110). An example of an alert-rule may be the detection of more than a predefined number of people in a zone covered by one of video sensors 12. Another example of an alert-rule may be the detection of a movement of an object larger than a predefined size from the right side to the left side of a zone covered by one of the sensors. Yet another example may be the detection of a particular face or a particular voice.

Each alert, sent by one of processing units 16 or 18 to control unit 10, may also be stored in database 30. The data stored may contain details about the alert such as the time of occurrence, the identifier of the sensor coupled to the processing unit providing the alert and the like.

Upon receiving an alert, control unit 20 may send a message to at least one of clients 40 notifying about the alert. Additionally or alternatively, control unit 20 may command application bank 24 to alter the applications installed in some of the processing units 16 and/or 18. Alternatively, control unit may directly command  
5 processing units 16 and/or 18 to activate or deactivate any application installed in the units (step 112). The new commands may be set according to predefined post-alert action-rules installed in control unit 20.

A non-limiting example of a post-alert action-rule may be: If one of video sensors 12 detects a movement, install face recognition application 28 in the processing  
10 unit 16, which is coupled to that sensor. Another example of a post-alert action-rule may be: If a particular person is identified by one of processing units 16, activate the compression application and record the video signal of the sensor 12 coupled to that processing unit. A third example may be: If one of audio sensors 14 identifies the voice of a particular person, install face recognition application to a specific processing unit 16  
15 coupled to video sensor 12 and start compression and recording of the video signal of that sensor.

The internal rules of control unit 20 may include the alteration of at least certain of the internal rules according to a time-based scheduler (not shown) stored in control unit 20.

20 Reference is now made to Figs. 4A and 4B, which are block diagrams of video-processing unit 16 of Fig. 1 according to some embodiments of the present invention. For clarity, Figs 4A and 4B and the description given hereinbelow refer only to video-processing units. However, it will be appreciated by persons skilled in the art that audio-processing units 18 may have similar structure.

Video-processing unit 16A may comprise an analog to digital (A/D) video signal converter 50 as illustrated in Fig 4A. A/D video converter 50 may receive analog video signals from one of video sensors 12 and to convert the analog signals into digital video signals.

5 Alternatively, video-processing unit 16B may comprise an Internet protocol (IP) to digital video signal converter 51 as illustrated in Fig 4B. Converter 51 may receive video signal over IP protocol from one of video sensors 12 and to extract video signals from the IP protocol.

10 Video-processing unit 16 may further comprise a processing module 52, an internal control unit 54, and a communication unit 56. Internal control unit 54 may receive applications from application bank 24 and may install the applications in processing module 52. Internal control unit 54 may further receive commands from control unit 20 and to alert control unit 20 when a condition corresponding to a rule is detected.

15 Processing module 52 may be a digital processor able to execute the applications installed by application bank 24. More than one application may be installed in video-processing unit 16. Processing unit 16 may further compress the audio and video signal and to transfer the compressed data to storage media 32 via communication unit 56. Processing module 52 may further transfer indexing data and the results of the  
20 applications to database 30 via communication unit 56. Non-limiting examples of communication unit 56 include a software interface, CTI interface, and an IP modem.

The following examples are now given, though by way of illustration only, to show certain aspects of some embodiments of the present invention without limiting its scope.

### EXAMPLE I

An operator commands control unit 20 via client 40:

- Install in all video-processing units a video compression application.
- Install at 08:00, in video-processing units coupled to video sensors #V1 - #V2 a face-recognition application and at 18:00 a motion detection application.
- Install in video-processing units coupled to video sensors #V11 - #V16 a people-counting application.
- Install in video-processing units coupled to video sensors #V17 - #V20 a motion detection application.
- Record for one minute the compressed video data received from any processing unit if a motion is detected or if the face-recognition application fails to identify a face.
- If more than 20 people are detected by video sensors #V11 - #V16, compress the video data until the number of people is less than 20.
- If a movement is detected by more than 30 video sensors within an hour, install people-counting application in video-processing units coupled to video sensors #V21- #V30.

### EXAMPLE II

Mr. X has to be located immediately.

An authorized operator commands control unit 20 via client 40 to add at least one rule regarding Mr. X

- Install in all video-processing units a face-recognition application.
- Install in all audio-processing units a voice-recognition application.

- Notify control unit when Mr. X is located.

### EXAMPLE III - OFF LINE INVESTIGATION

Calculating the number of people in the lobby at 08:00 - 08:30 and at 17:00 - 17:30,

5 Monday to Friday.

- An operator downloads a people-counting application to client 40.
- The operator requests playback of recorded video data from the video sensor installed in the lobby according to the required times.
- Client 40 executes the application and send a report to its display and/or printer

42.

While certain features of the invention have been illustrated and described herein, many modifications, substitutions, changes, and equivalents will now occur to those of ordinary skill in the art. It is, therefore, to be understood that the appended claims are intended to cover all such modifications and changes as fall within the true spirit of the invention.